基于改进混沌分区算法的模糊信息抽取 *

万福成

(中国民族语言文字信息技术教育部重点实验室，兰州 730000)

摘 要：在大数据环境下进行模糊信息挖掘抽取中受到数据之间的小扰动类间干扰的影响，导致信息抽取的特征聚类性不好，提出一种基于改进混沌分区算法的模糊信息抽取方法。对高维数据信息流进行分布式结构重组，以 Lorenz 混沌吸引子作为训练测试集进行大数据模糊信息抽取的自适应学习训练，采用相空间重构技术对大数据的混沌吸引子负载特征量进行自相关特征匹配处理，提取模糊信息的平均互信息特征量，结合关联规则模糊配对方法进行大数据混沌分区，实现模糊信息的优化聚类，根据数据聚类结果实现模糊信息准确抽取，对抽取的高维模糊信息进行特征压缩，降低计算开销。仿真结果表明，采用该方法进行大数据样本序列的模糊信息抽取的聚类性较好，抗类间扰动能力较强，模糊信息抽取的准确概率较高，在数据挖掘和特征提取中具有很好的应用价值。

关键词：大数据；混沌；分区算法；聚类；模糊信息抽取

中图分类号：TP391 doi: 10.3969/j.issn.1001-3695.2018.03.0209

Fuzzy information extraction based on improved chaotic partition algorithm

Wan Fucheng

(*Key Laboratory of National Language Intelligent Processing*, *Lanzhou* 730030, *China*)

Abstract: In the environment of big data, the interference between the small disturbances of the data affects the fuzzy information extraction, which leads to the poor clustering characteristics of information extraction. We propose a fuzzy information extraction method based on the improved chaotic partition algorithm. The high dimensional data information flow is reorganized with distributed structure, and the Lorenz chaotic attractor is used as the training test set for the adaptive learning training of big data fuzzy information extraction. We use the phase space reconstruction technique match big data's chaotic attractor load with autocorrelation feature matching and extract the average mutual information feature quantity of fuzzy information. Through realizing the optimal clustering of fuzzy information, the accurate extraction of fuzzy information is realized according to the result of data clustering, carries out the feature compression of the extracted high-dimensional fuzzy information, and reduces the computational overhead. The simulation results show that, using this method to extract fuzzy information from big data sample sequence has good clustering property, strong ability to resist inter-class disturbance, and high accurate probability of fuzzy information extraction. It has a good application value in data mining and feature extraction.

Key words: large data; chaos; partition algorithm; clustering; fuzzy information extraction

chinaXiv:201806.00116v1

0 引言

在大规模的数据中进行有效的数据区分和挖掘，如何利用和管理好大数据成为人们需要重点解决的问题，在大数据环境下，对模糊信息进行准确的特征分析和抽取，快速挖掘人们需要的数据信息，对数据进行整合，实现信息共享和准确链接传输，成为未来信息处理和大数据分析的重点研究课题，研究大数据环境下的模糊信息抽取方法，在云数据库设计、信息检索、语义分析以及信息融合等领域同样具有重要的应用价值[1]，相关的算法研究受到人们的极大重视。

对大数据环境下的模糊信息抽取方法研究是建立在对大数据聚类和信息分区基础上的，根据大数据的分类属性进行信息分区处理，提取大数据的聚类信息特征，采用特征分解和关联规则挖掘方法，实现对大数据的模糊信息抽取，常见的大数据模糊信息抽取方法有 HPCC (high performance computing cluster) 抽取方法，Roxie（HPCC data delivery engine）信息聚类特征抽取方法、基于频繁项挖掘的特征抽取方法、基于模糊 C-均值聚类的模糊数据信息抽取方法等[2,3]，通过提取模糊大数据的属性

收稿日期：2018-03-28；**修回日期：**2018-05-07 **基金项目：**国家自然科学基金资助项目（61762076）

作者简介：万福成（1985-），男，副教授，博士，主要研究方向为自然语言处理（wanfucheng@126.com）.

chinaXiv:201806.00116v1

分区特征。采用相应的聚类算法实现对大数据的模糊信息抽取，结合关联性的信息融合方法，提高模糊信息检测能力，根据上述原理，相关学者进行模糊信息抽取算法研究，取得了一定的研究成果，其中，文献[4]中提出一种基于类间闭频繁项集挖掘的混合大数据的相似度信息抽取方法，采用多层次的语义特征抽取方法进行大数据信息挖掘，根据用户的 Jarccad 相似度信息实现大数据模糊分区，提高信息的融合聚类能力，但该方法进行模糊信息抽取中存在计算开销过大和相似度特征分布稀疏性较强等；文献[5]中提出一种基于离散高斯随机检验的模糊大数据信息抽取方法，采用分段线性融合方法进行模糊信息的高斯随机信息特征提取，采用匹配滤波进行冗余信息过滤，提高大数据的模糊信息抽取的统计分析能力，该方法对模糊集的特征信息融合中受到类间干扰较大，容易导致模糊信息的错分和漏分。

针对上述问题，本文利用混沌的随机聚类性和抗扰动性，提出一种基于改进混沌分区算法的模糊信息抽取方法。首先构建大数据的模糊信息流模型，进行大数据的信息重组和相空间重构，然后利用混沌分区算法进行特征提取和数据聚类，实现大数据集中模糊信息优化抽取。最后进行实验分析，展示了本文方法在提高大数据模糊信息抽取准确性方面的优越性能。

1 模糊信息分布式结构重组及相空间重构

1.1 模糊信息分布式结构重组

为了实现对大数据中的模糊信息抽取，采用混沌分区方法进行关联规则挖掘，提取大数据中的模糊信息特征量，对大数据模糊信息时间序列混沌序列分析的第一步是进行相空间重构[6]，实现对模糊信息的分布式的结构重构和数据结构分析，假设待挖掘的大数据信息流观测时间序列 $\{x_1,x_2,\cdots x_N\}$，令 $x(n)$ 为一组非平稳的宽带时间序列，在 m 维分布式特征空间中进行模糊信息的结构映射，得到大数据的分布式重组结构式为

$$\boldsymbol{X}(n)=\{x(n),x(n+\tau),\cdots,x(n+(m-1)\tau)\}$$
$$n=1,2,\cdots,N \tag{1}$$

其中：τ 表示大数据在高维空间中的采样时间延迟，采用自适应的混沌训练方法进行特征融合，在高维特征空间中进行相轨迹演化分析[7]，得到重构的大数据模糊信息的分布轨迹为

$$X=[s_1,s_2,\cdots s_K]$$
$$=\begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_{1+\tau} & x_{2+\tau} & \cdots & x_{K+\tau} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1+(m-1)\tau} & x_{2+(m-1)\tau} & \cdots & x_{M+(m-1)\tau} \end{bmatrix} \tag{2}$$

其中：$K=N-(m-1)\tau$，表示的大数据搜索特征空间的嵌入维数，τ 为时延，m 为信息的语义本体属性的层数，$s_i=(x_i,x_{i+\tau},\cdots,x_{i+(m-1)\tau})^T$ 称为相空间的特征矢量集。在模糊信息分布式结构重组模型中，采用模糊集搜索方法确定 m 和时间延迟 τ。在高维相空间中进行模糊信息的高阶统计量分析和特征提取。

1.2 大数据相空间重构

在对高维数据信息流进行分布式结构重组的基础上，进行大数据相空间重构，以 Lorenz 混沌吸引子作为训练测试集进行大数据模糊信息抽取的自适应学习训练[8]，在相空间中的两个子空间分别为 S 和 Q，S 由模糊训练集的解向量 $\{s_1,s_2,\cdots,s_n\}$ 构成，满足 Lorenz 混沌吸引子的聚类属性的条件概率为 $P_s(s_i)$，$i=1,2,\cdots,n$，Q 由大数据模糊测试集的解向量 $\{q_1,q_2,\cdots,q_n\}$ 构成，对应的大数据模糊信息抽取的检测概率为 $P_q(q_j)$，$j=1,2,\cdots,n$。在混沌分区训练下，构建大数据模糊信息的相空间重构模型为

$$H(S)=-\sum_{i=1}^{n}P_s(s_i)\log_2 P_s(s_i) \tag{3}$$

$$H(Q)=-\sum_{i=1}^{n}P_q(q_j)\log_2 P_q(q_j) \tag{4}$$

其中：$P_s(s_i)$ 表示模糊信息的语义概念集 s_i 出现在混沌分区区域 S 的概率，类似地，$P_q(q_j)$ 表示模糊信息的本体特征概念集 q_j 出现在混沌分区区域 Q 的概率。在满足相似度一直的条件下，求相空间 S 中满足模糊信息聚类条件的平均互信息为

$$I(Q,S)=H(Q)-H(Q|S) \tag{5}$$

其中：

$$H(Q|s_i)=-\sum_{j}\left[p_{sq}(s_i,q_j)\big/p_s(s_i)\right]\log_2\left[p_{sq}(s_i,q_j)\big/p_s(s_i)\right] \tag{6}$$

表示满足条件 P 的分类属性集合，采用平均互信息融合方法。根据上述相空间重构结构，提取模糊信息的高维特征量，结合混沌分区算法进行模糊聚类，提高模糊信息挖掘的聚类性[9]。

2 模糊信息抽取算法实现

2.1 改进的混沌分区算法

在对高维数据信息流进行分布式结构重组和相空间重构预处理的基础上，进行大数据的模糊信息抽取算法优化设计，本文提出一种基于改进混沌分区算法的模糊信息抽取方法。以 Lorenz 混沌吸引子作为训练测试集，进行大数据的模糊属性分区，给出 Lorenz 混沌吸引子表达式：

$$\begin{cases} dx/dt=-\sigma x+\sigma y \\ dy/dt=-xz+rx-y \\ dz/dt=xy-bz \end{cases} \tag{7}$$

式中各量均为约化后的无量纲量，t 为表示对模糊数据的采样时间；x，y，z 表示为 Lorenz 混沌吸引子的分区变量；σ，r，b 为模糊约束参量。采用四阶 Runge-Kutta 法进行模糊信息的混沌分区，对于模糊信息的特征矢量 $\boldsymbol{X}_n$，在混沌分区的有限域中选择近邻轨迹向量 $\boldsymbol{X}_{\eta(n)}$，令 R_{mn} 为 $\boldsymbol{X}_n$ 与 $\boldsymbol{X}_{\eta(n)}$ 的距离，得到大数据混沌分区的聚类中心欧氏距离为

chinaXiv:201806.00116v1

$$R_{mn}=\left\|\boldsymbol{X}_{\eta(n)}-\boldsymbol{X}_n\right\|_2^{(m)}=\min_{j=N_0,\cdots,N,j\neq n}\left\|\boldsymbol{X}_n-\boldsymbol{X}_j\right\|_2=\sqrt{\sum_{l=0}^{m-1}(x_{\eta(n)+l\tau}-x_{n+l\tau})^2} \tag{8}$$

在上述重构的相空间中进行自适应学习，将相空间的嵌入维数从 m 增加到 $m+1$ 时，得到模糊信息抽取的混沌分区的聚类中心优化值表示为

$$R_{(m+1)n}=\left\|\boldsymbol{X}_{\eta(n)}-\boldsymbol{X}_n\right\|_2^{(m+1)}=\sqrt{\sum_{l=0}^{m}(x_{\eta(n)+l\tau}-x_{n+l\tau})^2}=\sqrt{\left[\left\|\boldsymbol{X}_{\eta(n)}-\boldsymbol{X}_n\right\|_2^{(m)}\right]^2+\left[x_{\eta(n)+m\tau}-x_{n+m\tau}\right]^2} \tag{9}$$

若 $R_{(m+1)n}$ 比 R_{mn} 大很多，表示模糊信息的类间属性扰动互不影响，可以作为模糊聚类中心，根据这一特性，得到的 $\boldsymbol{X}_{\eta(n)}$ 为 $\boldsymbol{X}_n$ 的最近邻点。构建判决统计量和判决准确，在混沌分区中进行模糊信息抽取的阈值判断，得到判决准则为

准则 1

$$\sqrt{\frac{R_{(m+1)n}{}^2-R_{mn}{}^2}{R_{(m+1)n}{}^2}}=\frac{\left|x_{\eta(n)+m\tau}-x_{n+m\tau}\right|}{R_{(m+1)n}}\geq R_{tol} \tag{10}$$

准则 2

$$\frac{R_{(m+1)n}}{\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[x_k-\frac{1}{N}\sum_{k=1}^{N}x_k\right]^2}}>A_{tol} \tag{11}$$

其中：R_{tol} 为 Lorenz 混沌吸引子的分区阈值，根据经验值，R_{tol} 可取 15，A_{tol} 表示模糊信息混沌分区的判决门限，取 $A_{tol}=2$。

根据上述分析，实现对大数据环境下的模糊信息混沌分区，结合判决准则和判决统计量进行大数据模糊信息抽取的自适应学习训练，并进行特征提取和信息抽取[10]。

2.2 信息抽取和特征压缩

假设模糊信息的分布时间序列 $\{X_n\},n=1,2,\cdots,N$，代表原始待分区的大数据特征分布集，在混沌分区处理下，得到混沌分区后的特征分布 $X_N=X_n+\eta$，其中 η 为观测噪声。在 d 个大数据的分布信源中，采用相空间重构技术对大数据的混沌吸引子负载特征量进行自相关特征匹配处理[11]，得到特征匹配输出：

$$X_n=\{X_n,X_{n-\tau},X_{n-2\tau},\cdots,X_{n-(d-1)\tau}\} \tag{12}$$

令 $R_{d\times L}$ 是 $d\times L$ 的矩阵，在混沌分区的闭频繁项区域中提取模糊信息的平均互信息特征量为

$$R_1=\{X_1,X_2,X_3,\cdots,X_d\}^T \tag{13}$$

模糊信息的关联规则向量集为

$$R_1^TR_1=\{X_1,X_2,\cdots,X_m\}\{X_1,X_2,\cdots,X_m\}^T \tag{14}$$

采用奇异分解方法对模糊信息特征值进行特征分解，得

$$R_1^TR_1=V_1\sum{}_1V_1^T \tag{15}$$

对 $L+1$ 到 $2L$ 维的混沌分区大数据，根据上述方法类推，得到模糊信息抽取的输出特征值为

$$R_2^TR_2=V_2\sum{}_2V_2^T \tag{16}$$

$$R_2=\{X_{d+1},X_{d+2},\cdots X_{d+m}\}^T \tag{17}$$

$$R_2^TR_2=\{X_{d+1},X_{d+2},\cdots X_{d+m}\}\{X_{d+1},X_{d+2},\cdots X_{d+m}\}^T \tag{18}$$

其中：模糊信息的测试集 $V=[V_1,V_2,\cdots,V_m]\in R^{m\times m}$ 是正交的，即 $VV^T=I_M$，$\sum=diag(\sigma_1,\sigma_2,\cdots,\sigma_m)\in R^{m\times m}$，混沌分区后的关联规则模糊特征向量集 R^TR 满足类间平衡性，得到抽取的特征值聚类融合度排序为

$$\sigma_1>\sigma_2>\sigma_3>\cdots>\sigma_{s+1}>\sigma_m \tag{19}$$

分析可见，采用上述方法抽取的某信息维数较高，需要进行特征压缩，特征压缩的步骤描述为：

a) 计算混沌分区区域中的 l 维模糊信息特征向量 $\vec{X}(l,n_i)$，根据分区的聚类属性值得到模糊信息分区离散度 $\hat{S}_w$，计算模糊信息的 l 个特征值 $\lambda_1,\lambda_2,\cdots,\lambda_l$ 和对对应的高维特征量 $Y=[y_1,y_2,\cdots,y_l]$。$\hat{S}_w$ 为

$$\hat{S}_w=\sum_{i=1}^{c}p_i\frac{1}{n_i}\sum_{k=1}^{n_i}\left[\left(\vec{X}_k^{(i)}-\vec{m}_i\right)\left(\vec{X}_k^{(i)}-\vec{m}_i\right)^T\right] \tag{20}$$

b) 根据混沌分区的吸引子的中心矢量，得到模糊聚类的离散度矩阵 $S_b=\sum_{i=1}^{c}p_i\left(\vec{m}_i-\vec{m}\right)\left(\vec{m}_i-\vec{m}\right)^T$，其中 $\vec{m}=\sum_{i=1}^{c}p_i\vec{m}_i$ 的加权平均；

c) 采用判决统计回归分析方法，在高维空间中进行特征压缩，输出特征量 $J\left(\vec{X}_j\right)=\frac{y_j^TS_by_j}{\lambda_j}$，$j=1,2,\cdots,l$，采用 K-L 变换进行特征排序，得到特征排列式 $J\left(\vec{X}_1\right)\geq J\left(\vec{X}_2\right)\geq\cdots\geq J\left(\vec{X}_l\right)$；

d) 输出 d 维的模糊信息抽取特征量，抽取前 d 个特征向量 $J\left(\vec{X}_j\right)$ 作为训练集，输出 y_j，$j=1,2,\cdots,d$，得到模糊变换加权矩阵 $W=\left[y_1,y_2,\cdots,y_d\right]$；

e) 输出特征压缩后的信息抽取结果为 $\vec{X}^*=W^T\vec{X}$。

通过上述处理，在混沌分区后输出的模糊信息特征量由 l 降到 d 维，从而降低了计算开销。

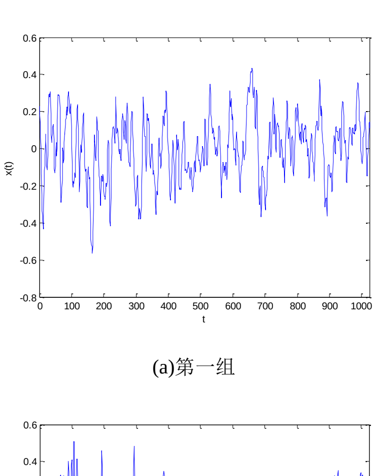
3 仿真实验与结果分析

为了验证本文方法在实现大数据集下的模糊信息抽取的应用性能，进行仿真实验，实验采用 Visual C++进行算法编译，采用 MATLAB 进行数据处理的编程设计，在 MapReduce 编程

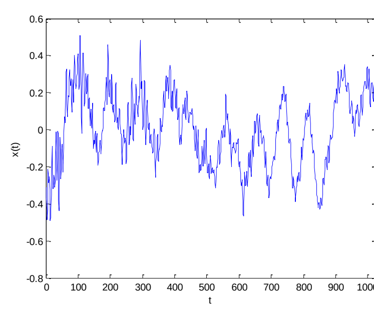
ChinaXiv合作期刊

chinaXiv:201806.00116v1

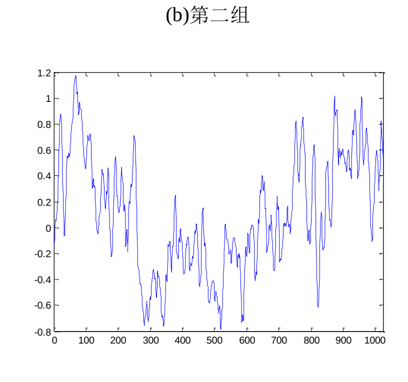
框架进行数据抽取和信息聚类分析，大数据集来自于云平台 Hadoop，大数据处理工具组件为 Thor（HPCC data refinery cluster），大数据测试样本集为 OAEI (ontology alignment evaluation initiative)，该大数据库存储了多版本的语义数据，作为本实验的数据测试集能具有较好的信息覆盖能力。数据采样的时间间隔为 2.4s，测试集的规模为 2400Gbit,大数据训练集的采样样本长度为 1024，Lorenz 混沌吸引子的初始参数值设定为 $[x,y,z]=[-1,0,1]$，$[\sigma,r,b]=[16,45.92,4.0]$，混沌分区迭代的训练步长为 $h=0.01$，大数据聚类的信息分布区间为 $[0,1000]$，在信息抽取中的干扰强度为-10dB，根据上述仿真环境和参数设定，进行大数据模糊信息抽取仿真实验，取三组测试样本，得到时域波形如图 1 所示。



(a)第一组



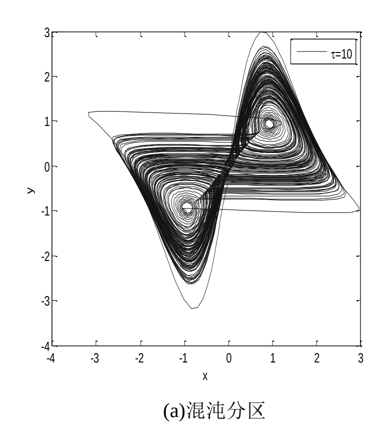
(b)第二组



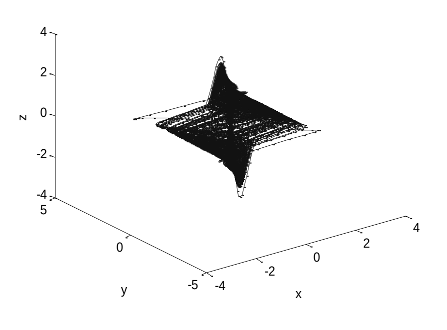
(c)第三组

图 1 测试样本时域波形

以图 1 的数据样本为研究对象，采用本文方法进行混沌分区聚类和信息抽取，得到在最佳的相空间嵌入维和时间延迟下的各组样本的混沌分区和模糊信息抽取结果如图 2~4 所示。

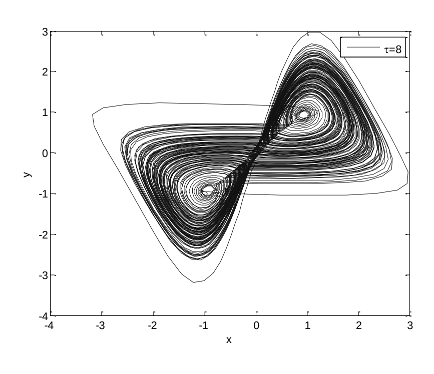


(a)混沌分区

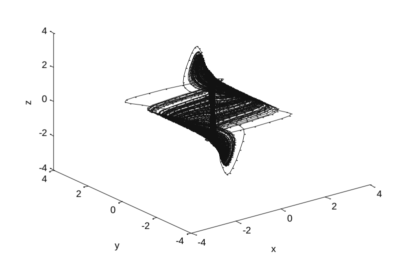


(b)模糊信息抽取

图 2 第一组样本的混沌分区和信息抽取结果



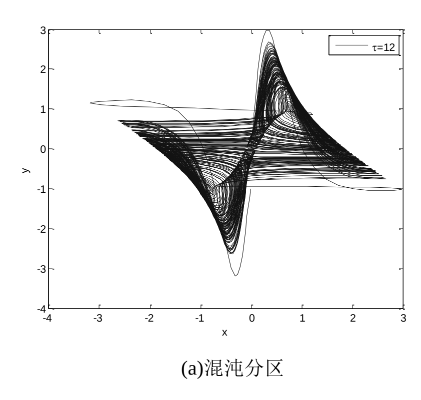
(a)混沌分区



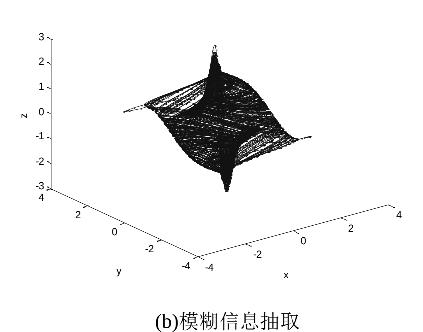
(b)模糊信息抽取

图 3 第二组样本的混沌分区和信息抽取结果

ChinaXiv合作期刊



(a)混沌分区



(b)模糊信息抽取

图 4 第三组样本的混沌分区和信息抽取结果

图 2 对应第一组数据样本，该组样本的时域波形波动较为稳定，表明该组数据的信息覆盖能力较强，而由图 2 可以看出，利用本文方法进行混沌分区聚类和信息抽取，得到在最佳的相空间嵌入维和时间延迟下的混沌分区和模糊信息抽取结果清晰，属性聚类性较好，表明所提方法对大数据信息可以实现准确信息抽取，这是由于所提方法在数据环境下，结合了判决准则和判决统计量进行了大数据模糊信息抽取和自适应学习训练，并进行特征提取，由此使其更加适用。图 3 对应的第二组数据样本，该组样本数据整体波动次数少，幅度较为一致，表明其数据分布较为分散、数据维数较高，本身聚类效果较差。而使用所提方法对其进行混沌分区和模糊信息抽取后，其混沌分区的属性聚类性较好，信息抽取的类间聚敛性较高，这是由于所提方法利用了自适应混沌训练方法进行特征融合，在高维特征空间中进行了相轨迹演化分析，重构了大数据模糊信息，从而提高了其混沌分区与模糊信息抽取能力。图 3 对应第三组数据样本，该组数据样本的时域波形波动无规律，且幅度较大，表明该组数据受到的类间干扰较大，而使用本文所提方法仍可以保证较好的混沌分区和模糊信息提取性能，表明所提方法的抗类间干扰能力较强，这主要是由于本文所提方法在对维数较高的信息抽取时，同时进行了特征压缩，降维处理有效降低了处理数据样本受干扰程度，改善其性能。

分析上述结果得知，采用本文方法进行大数据信息的混沌分区的属性聚类性较好，信息抽取的类间聚敛性较高，抗类间扰动能力较强，具有很好的模糊信息抽取能力。为了对比性能，采用不同方法，对各组样本进行模糊信息抽取，分析抽取精度和计算开销等参数指标，得到性能对比结果见表 1 和表 2，分析表中结果得知，本文方法进行模糊信息抽取时间开销较小，抽取的准确度较高，性能优于传统方法。

表 1 模糊信息抽取的时间开销性能对比(单位:s)

样本	本文方法	FCM	K-means
第一组	0.24	1.43	1.25
第二组	0.31	2.78	3.21
第三组	0.52	2.96	4.32

表 2 抽取精度性能对比(%)

样本	本文方法	FCM	K-means
第一组	98.23	92.34	89.32
第二组	99.34	94.93	92.45
第三组	99.98	95.65	94.31

4 结束语

在大数据环境下，对模糊信息进行准确的特征分析和抽取，快速挖掘人们需要的数据信息，对数据进行整合，实现信息共享和准确链接传输，本文利用混沌的随机聚类性和抗扰动性，提出一种基于改进混沌分区算法的模糊信息抽取方法。采用 Lorenz 混沌吸引子进行大数据混沌分区，在重构的相空间中实现模糊信息抽取，并对抽取的高维数据进行特征降维。研究得知，本文方法能提高模糊信息抽取的精度，计算开销较小，抗类间干扰能力较强，性能优越。

参考文献：

[1] 梁聪刚，王鸿章. 微分进化算法的优化研究及其在聚类分析中的应用 [J]. 现代电子技术，2016, 39 (13): 103-107. (Ling Conggang, Wang Hongzhang. Optimization research on differential evolution algorithm and its application in clustering analysis [J]. Modern Electronic Technology, 2016, 39 (13): 103-107.)

[2] 米捷，张鹏，于海鹏. 粒子群差分扰动优化的聚类算法研究 [J]. 河南工程学院学报，2016, 28 (1): 63-68. (Mi Jie, Zhang Peng, Yu Haipeng. Large data clustering algorithm based on particle swarm differential perturbation optimization [J]. Journal of Henan University of Engineering (Natural Science Edition) , 2016, 28 (1): 63-68.)

[3] 邢淑凝，刘方爱，赵晓晖. 基于聚类划分的高效用模式并行挖掘算法 [J]. 计算机应用，2016, 36 (8): 2202-2206. (Xing Shuning, Liu Fang'ai, Zhao Xiaohui. Parallel high utility pattern mining algorithm based on cluster partition [J]. Journal of Computer Applications, 2016, 36 (8): 2202-2206.)

[4] Palomares I, Martinez L, Herrera F. A consensus model to detect and manage non-cooperative behaviors in large scale group decision making [J]. IEEE Trans on Fuzzy System, 2014, 22 (3): 516-530.

[5] 李梓杨，于炯，卞琛，等. 基于负载感知的数据流动态负载均衡策略 [J]. 计算机应用，2017, 37 (10): 2760-2766. (Li Ziyang, Yu Jiong, Bian Chen, *et al.* Dynamic data stream load balancing strategy based on load awareness [J]. Journal of Computer Applications, 2017, 37 (10): 2760-2766.

chinaXiv:201806.00116v1

[6] 孙力娟，陈小东，韩 崇，等. 一种新的数据流模糊聚类方法 [J]. 电子与信息学报, 2015, 37 (7): 1620-1625. (Sun Lijuan, Chen Xiaodong, Han Chong, *et al.* New Fuzzy-Clustering Algorithm for Data Stream [J]. JEIT, 2015, 37 (7): 1620-1625.)

[7] 邢长征，刘剑. 基于近邻传播与密度相融合的进化数据流聚类算法 [J]. 计算机应用，2015, 35 (7): 1927-1932. XING Changzheng, LIU Jian. Evolutionary data stream clustering algorithm based on integration of affinity propagation and density. Journal of Computer Applications, 2015, 35 (7): 1927-1932.)

[8] 侯森，罗兴国，宋克. 基于信息源聚类的最大熵加权信任分析算法 [J]. 电子学报, 2015, 43 (5): 993-999. (Hou Sen, Luo Xingguo, Song Ke. A maximum entropy weighted trust-analysis algorithm based on sources clustering [J]. Chinese Journal of Electronics, 2015, 43 (5): 993-999.)

[9] 毕安琪，王士同. 基于 Kullback-Leiber 距离的迁移仿射聚类算法 [J]. 电子与信息学报, 2016, 38 (8): 2076-2084. (Bi Anqi, Wang Shitong. Transfer affinity propagation clustering algorithm based on Kullback-Leiber Distance [J]. JEIT, 2016, 38 (8): 2076-2084.)

[10] 刘俊，刘瑜，何友，等. 杂波环境下基于全邻模糊聚类的联合概率数据互联算法 [J]. 电子与信息学报, 2016, 38 (6): 1438-1445. (Liu Jun, Liu Yu, He You, *et al.* Joint Probabilistic data association algorithm based on all-neighbor fuzzy clustering in clutter [J]. JEIT, 2016, 38 (6): 1438-1445.)

[11] 吴鸿华，穆勇，屈忠锋，等. 基于面板数据的接近性和相似性关联度模型 [J]. 控制与决策, 2016, 31 (3): 555-558. (Wu Honghua, Mu Yong, Qu Zhongfeng, *et al.* Similarity and nearness relational degree based on panel data [J]. Control and Decision, 2016, 31 (3): 555-558.)

chinaXiv:201806.00116v1